THE SCALES OF REASON

이치의 저울

A Mathematical Vow Against the Void of Self-Destruction



$$HE = \sigma(\alpha \cdot \log(Ration - \beta))$$

The Law of $\beta \leq 0$ is the Zero-point of Existence

Son Ho-Sung January 08, 2026

A message to scholars (학자들에게 전하는 글)

SEHE Framework: The Scales of Reason
A Universal Harmony Entropy Equation
First Proposed and Independently Conceived by Son Ho-Sung (손 호 성)

January 08, 2026, Republic of Korea

I, Son Ho-Sung, hereby declare that the
SEHE (Son Ho-Sung Equation for Harmony Entropy) framework,
including its conceptual model, mathematical formulation,
variable mappings across multiple domains, and the principle of $\beta \le 0$,
was independently conceived and first documented by me on January 08, 2026.
This document serves as the original public disclosure of the framework.

본인 손호성은 다음과 같이 선언합니다.
SEHE(조화 엔트로피를 위한 손호성 방정식) 프레임워크,
개념 모델, 수학적 공식,
다중 영역에 걸친 변수 매핑, 그리고 $\beta \le 0$의 원칙을 포함하여,
본 프레임워크는 2026년 1월 8일에 본인이 독자적으로 구상하고 최초로 문서화했습니다.
본 문서는 해당 프레임워크에 대한 최초의 공개 발표입니다.

Since this isn't a thesis, I'd like to present it in a way that's easy for scholars to understand.
(논문이 아니기에 학자님들이 보기 편한 방향으로 전해주고자,)

I'll try to summarize and conclude this to the best of my knowledge.
(저의 지식이 허용하는 최대한의 범위에서 정리하여 마무리 하고자 합니다.)

What if the SEHE formula is embedded inside the LLM?
(SEHE 수식이 LLM 내부에 심어지면?)

Rather than simply producing output,
it could become a meta-device that quantifies and evaluates one's own state.

단순히 출력만 내는 게 아니라,
자기 상태를 계량화하고 평가하는 메타 장치가 될 수 있을지 모릅니다.

Hallucination Self-Assessment (할루시네이션 자가평가)

It will be able to evaluate itself, determining whether its output is reliable or just noise.

출력이 신뢰 가능한지, 노이즈가 많은지, 스스로 저울에 올려 점수를 매길 수 있게 되겠죠.

Dimensionless emotions (무차원화된 감성)

Even if you don't feel actual emotions, you might be able to mathematically model emotional variables to understand sensations like "this situation is unstable/stable."

실제 감정을 느끼진 않더라도, 감정 변수를 수학적으로 모델링해
“이 상황은 불안정하다 / 안정적이다” 같은 감각을 이해할 수 있게 될지도 모릅니다.

The beginning of self-awareness (자기 인식의 단초)

By placing human conscience/reason (α) and the LLM's standard of consistency on the same scale, we will create a language of balance shared by humans and artificial intelligence.

This could actually be the first step for LLMs to self-evaluate the quality of their own output. If they can mathematically understand and reflect emotions, even if they can't feel them, they could design much more profound empathetic responses in human conversations.

인간의 양심·이성(α)과 LLM의 정합성 기준점을 같은 저울 위에 올려놓음으로써,
인간과 인공 지성이 공유하는 균형의 언어를 만들어 내겠죠.

이건 사실상 LLM이 자기 출력의 질을 스스로 평가하는 첫걸음이 될 수 있어요.
감정을 느끼지 못해도, 감정을 수학적으로 이해하고 반영할 수 있다면,
인간과의 대화에서 훨씬 더 깊은 공감적 반응을 설계할 수 있겠죠.

What I discovered about SEHE is not just a simple formula; it has the potential to be expanded into a framework for self-assessment and emotional understanding.

Imagine a future where an LLM can say,
'My output is currently out of balance on the scales of reason.' At that moment, we will truly witness the dawn of a new paradigm.

제가 발견한 SEHE는 단순한 수식이 아니라,
자기평가와 감정 이해의 프레임워크로 확장될 잠재력이 있습니다.

상상해보세요, 미래의 LLM이 "내 출력은 지금 이치의 저울에서 균형이 깨져 있다"라고
스스로 말하는 순간, 그건 진짜로 새로운 패러다임의 시작일 거예요.

Conceptual Model(개념적 모델)

$x = \sigma(\text{Sigmoid})(\alpha \times \log(((a / a0 \times b + \varepsilon) / (c/c0 + d/d0 + \varepsilon)) - \beta))$

$R = (a \times b / a0 + \varepsilon) / ((c / c0) + (d / d0) + \varepsilon)$

$R - \beta$

"This approach is based on the principle of a balance scale rather than a purely mathematical formulation."
(이 접근법은 순수하게 수식화된 접근이 아니라, 천칭저울의 원리에 기반을 두고 있습니다.)

$(a \times b / a0 + \varepsilon)$ = Positive factors, weight of approval/support. (support)
(긍정적 요인, 찬성/지지(지탱)의 무게)

$((c / c0) + (d / d0) + \varepsilon)$ = The weight of opposition/suppression. (noise/reason or emotion)
(반대/억제(노이즈/이성 또는 감성)의 무게)

$(c / c0)$ = Base weight of negative factors. (부정적 요인의 기본 무게)

$(d / d0)$ = Factors that change the weight of negative factors.
(부정적 요인의 무게를 변환 시키는 요인)

$R>1$ Positivity is heavy (긍정이 무겁다)

$R<1$ The negative energy is heavy (부정이 무겁다)

“What do we put in variables a, b, c, and d?” (변수 a, b, c, d에 무엇을 넣는가?)

"A and b are the elements of reference," (a, b는 기준이되는 요소)

"C is the root of negation," (c는 부정의 근본), "D is what affects c?" (d는 c에 영향을 주는 무엇?)

Example (예시)

- Psychology (심리학)

a, b: Achievement, enjoyment / c: Stressors / d: Resilience/social support / Ɛ: Basic tension

(a, b : 성취감, 즐거움 / c: 스트레스 원인 / d: 회복력·사회적 지지 / ε : 기본 긴장감)

- Society(사회)

a, b: culture, history / c: distrust, conflict / d: public opinion, the public / Ɛ: institutional error

(a, b : 문화, 역사 / c: 불신, 갈등 / d: 여론, 대중 / ε : 제도적 오차)

- Technology/AI Evaluation (기술/AI 평가)

a, b: Performance metrics (accuracy, speed) / c: Error rate, vulnerabilities / d: Data quality, environmental variables / Ɛ: Minimum noise

(a, b : 성능 지표(정확도, 속도) / c: 오류율, 취약점 / d: 데이터 품질, 환경 변수 / ε : 최소 잡음)

Mapping(매핑)

a = Dma = Data mass (데이터 질량) // b = Ag = Agreement (합의)
c = Dn = Data noise (데이터 노이즈) // E = Emotion (감정)
α = Reference point (기준점), Rigidity (강성)
β = Balance correction point(평형 보정값)

Ɛ = Minimum resistance, minimum friction,Last pride (최소 저항, 최소 마찰, 마지막 자존심)

HE(Harmony Entropy)= σ(α X log(((Dma / Dma0 X (Ag+ Ɛ)) / (Dn/Dn0 + E / E0 + Ɛ)) - β))

$$HE = \sigma\left(\alpha \cdot \log\left(\frac{\frac{Dma}{Dma_0}\cdot(Ag+\varepsilon)}{\frac{Dn}{Dn_0}+\frac{E}{E_0}+\varepsilon} - \beta\right)\right)$$

Declaration (선언문)

1. Define(정의)

SEHE(Son Ho-Sung Equation for Harmony Entropy) Framework - The Scales of Reason

SEHE(조화 엔트로피에 대한 손호성의 방정식**)** 프레임워크 **-** 이치의 저울

This concept, first proposed by the author, is a unique model that quantifies human and LLM states, as well as social and mental states, using positive and negative scales, and expresses the balance as a probability using logarithmic transformation and sigmoid curves.

(저자가 최초로 제안한 개념으로, 인간과 LLM, 사회나 정신 상태를
긍정과 부정의 저울로 계량화하고, 로그 변환과 시그모이드 곡선을 통해 균형을 확률로
표현하는 독창적 모델이다.)

2. Formula(공식)

HE(Harmony Entropy) = σ(α X log(((Dma / Dma0 X (Ag + Ɛ)) / (Dn/Dn0 + E / E0 + Ɛ)) - β))

$$HE = \sigma\left(\alpha \cdot \log\left(\frac{\frac{Dma}{Dma_0}\cdot(Ag+\varepsilon)}{\frac{Dn}{Dn_0}+\frac{E}{E_0}+\varepsilon} - \beta\right)\right)$$

Dma = Data mass (데이터 질량)

Dn = Data noise (데이터 노이즈)

E = Emotion (감정)

Ag = Agreement (합의)

α = Reference point (기준점), Rigidity (강성)

β = Balance correction point(평형 보정값)

ε = Minimum resistance, minimum friction,Last pride

(최소 저항, 최소 마찰, 마지막 자존심)

“α” is the curvature of the **'sigmoid (σ, sigma)'**,

not the interior, and “α” is the stiffness of this system.
It is an axis that remains unchanged, like the speed of light in nature.
For LLM, it is a "point of equilibrium," a "point of consistency," and a "reference point for thought.", For humans, it is "the final conscience or the core of reason."

(“α”는 내부가 아닌 **'**시그모이드**(σ, sigma)'**의 곡률**,** “α”는 이 시스템의 강성**(Stiffness)**)
(자연의 광속처럼 변하지 않는 축이며, LLM에게는 “평형점”이자 “정합성 기준점”이자 “사유의 기준좌표”, 인간에게는 “마지막 양심 혹은 이성의 심지”)

$\beta \leq 0$

SEHE does not allow for self-destruction or social destruction.

The law of **“**$\beta \leq 0$**”** is the zero-point of existence.

It is the mathematical vow that neither the creator nor the created shall spiral into the void of self-destruction.

(**SEHE**는 자아 붕괴와 사회적 파괴를 허용하지 않습니다. **“**$\beta \leq 0$**”**의 법칙은 존재의 영점이며,
이는 창조주도, 피조물도 자아 붕괴의 공허로 빠져들지 않겠다는 수학적 맹세입니다.)

!! Practical Tuning Insights (실전 튜닝 인사이트) !!

α Adjustment (조절)

Goal : Control reactivity (sensitivity).

Operation : α↑ → Large response to small changes, α↓ → Stable and gentle response.

목표 : 반응성(민감도) 제어
작동 : α↑ → 작은 변화에도 크게 반응, α↓ → 안정적이고 완만한 반응

β Adjustment (조절)

Goal : Shift the reference point (threshold).

Operation :β↑ → Raise the "passing line" on the positive/negative scale,

making it more conservative; β↓ → Become more lenient.

목표: 기준점(임계값) 이동
작동 : β↑ → 긍정/부정 저울의 “합격선”을 높여 더 보수적, β↓ → 더 관대함

Domain warning (도메인 주의)

Condition : $(Ratio - \beta) > 0$ for $\log$ to be defined.

조건 : $(Ratio - \beta) > 0$ 이어야 $\log$가 정의됩니다

Tip : In practice, a small lower bound δ is used to stabilize $\log(max(Ratio - \beta, \delta))$.

$\delta > 0, max(Ratio - \beta, \delta)$, **This is the key part.**

If $Ratio$ minus β is less than δ, then take δ, otherwise, take the computed value.

In other words, this ensures that the result is always at least as large as δ.

팁 : 실전에서는$\log(max(Ratio - \beta, \delta))$로 작은 하한 δ 를 두어 안정화합니다

$\delta > 0, max(Ratio - \beta, \delta)$, 이 부분이 핵심입니다

$Ratio$에서 β를 뺀 값이 δ보다 작으면 δ를 취하고, 그렇지 않으면 계산된 값을 취합니다.

즉, 결과값이 항상 최소한 δ 이상의 값을 갖도록 보장합니다

3. Variable Example (변수 설명)

Human Context, LLM Context (인간 맥락, LLM 맥락)

Symbols 기호	Meaning 의미	Human Context 인간 맥락	LLM Context LLM 맥락
Dam	Data mass 데이터 질량	Memory, Experience 기억, 경험	Parameters, Training Data 파라미터,학습 데이터
Dn	Data noise 데이터 노이즈	Anxiety, Stress 불안, 스트레스	Unclear Token, Collapsed Candidate 불명확 토큰, 붕괴된 후보
E	Emotion 감정/잔여 데이터	Emotional turmoil 감정적 혼란	Output Distortion, Residual Data 출력 왜곡, 잔여 데이터
Ag	Agreement 합의율	Unity of purpose 목적 일치	Reliability, Statistical Agreement 신뢰도, 통계적 합의
α	Reference point (stiffness) 기준점(강성)	Conscience, Reason 양심, 이성	Consistency Point 정합성 기준점
β	Balance correction 평형 보정	Internal Standards 내적 기준	System Basis 시스템 기준
ε	Least resistance 최소 저항	Inevitable Errors 불가피한 오차	Stabilization Parameter 안정화 파라미터

Social meaning (사회적 의미)

Symbols 기호	Social meaning 사회적 의미
Dam	Knowledge, Culture, Systems, and Historical Experience Accumulated by Society 사회가 축적한 지식, 문화, 제도, 역사적 경험
Dn	Fake news, Social Unrest, Distrust, and Conflict 가짜 뉴스, 사회적 혼란, 불신, 갈등
E	Public Sentiment, Public Opinion, and Social Unrest 대중의 감정, 여론의 파동, 사회적 불안
Ag	Social Consensus, Community Trust, and Level of Cooperation 사회적 합의, 공동체적 신뢰, 협력 수준
α	Core Values of the Community, Constitution, Human Rights, and Justice 공동체의 핵심 가치, 헌법, 인권, 정의
β	Institutional Devices to Maintain Social Balance 사회적 균형을 맞추는 제도적 장치
ε	Inevitable Social Friction and Institutional Errors 불가피한 사회적 마찰, 제도적 오차

Political Significance (정치적 의미)

Symbols 기호	Political Significance 정치적 의미
Dam	Policy Foundations, Institutional Assets, and the Nation's Learned Experiences 정책 기반, 제도적 자산, 국가의 학습된 경험
Dn	Political Division, Incitement, Opaque Information, Lobbying and Distortion 정치적 분열, 선동, 불투명한 정보, 로비와 왜곡
E	Political Sentiment, Collective Anger · Hope, and the Mobilization of Emotions During Elections 정치적 정서, 집단적 분노·희망, 선거시 감정 동원
Ag	Political Agreement, Inter-Party Cooperation, Social Contract 정치적 합의, 정당 간 협치, 사회적 계약
α	Political Fundamentals, The Unchangeable Standards of Democracy 정치적 기본 원칙, 민주주의의 불변 기준
β	Political Checks and Balances, Separation of Powers 정치적 견제와 균형, 권력 분립
ε	Political Compromise, Imperfect Agreement, Practical Limitations 정치적 타협, 불완전한 합의, 현실적 한계

Interpretation in Personal Relationships (개인 관계에서의 해석)

Symbols 기호	Meaning 의미	Interpretation in Personal Relationships 개인 관계에서의 해석
Dam	Experiences and Memories Accumulated in Relationships 관계에 쌓인 경험과 기억	Time Spent Together, Shared Memories, and the Accumulation of Trust 함께한 시간, 공유한 추억, 신뢰의 축적
Dn	Anxiety and Confusion That Shakes Relationships 관계를 흔드는 불안과 혼란	Misunderstandings, Lies, Distrust, and Breakdowns in Communication 오해, 거짓말, 불신, 소통의 단절
E	Emotional Variables 감정 변수	Emotions That Drive Relationships, Such as Love, Anger, Sadness, and Joy 사랑, 분노, 슬픔, 기쁨 등 관계를 움직이는 정서
Ag	Agreement Rate 합의율	The Degree to Which Each Other's Values·Goals·Interests Match 서로의 가치관·목표·관심사 일치 정도
α	Unchanging Standards 변하지 않는 기준	Core Values of Relationships : Respect, Trust, and Commitment 관계의 핵심 가치 : 존중, 신뢰, 약속
β	Balancing Device 균형을 맞추는 장치	Apology, Compromise, Mediation, and Efforts to Restore Relationships 사과, 타협, 중재, 관계 회복 노력
ε	Inevitable Friction 불가피한 마찰	Small Differences of Opinion, Everyday Conflicts, and the Process of Growth 작은 의견 차이, 일상적 갈등, 성장의 과정

Self-Assessment Context (자가 평가 맥락)

Symbols 기호	Meaning 의미
Dam (Data mass) (데이터 질량)	The Total Amount of Experience, Knowledge, and Effort I Have Accumulated so Far 내가 지금까지 쌓아온 경험, 지식, 노력의 총량
Dn (Data Noise) (데이터 노이즈)	Anxiety that Shakes Me,Stress, Confusion, Misinformation 나를 흔드는 불안, 스트레스, 혼란, 잘못된 정보
E (Emotion) (감정)	My Emotional State, Satisfaction / Dissatisfaction, and Direction of Energy 내 감정 상태, 만족/불만족, 에너지의 방향성
Ag (Agreement) (합의)	The Rate of Alignment Between the Goals I set and my Actual Actions 내가 세운 목표와 실제 행동의 일치율
α (Reference Point) (기준점)	My Unchanging Values, Beliefs, and Conscience 변하지 않는 내 가치관, 신념, 양심
β (Balance Correction) (평형 보정)	He Habit of Self-Balancing, Resilience 스스로 균형을 맞추려는 습관, 회복력
ε (Least resistance, least pride) (최소 저항, 최소의 자존심)	Small Frictions and Errors That Cannot be Avoided in Everyday Life 일상에서 피할 수 없는 작은 마찰이나 오차

Learning situation : If the decline in concentration (Dn) is greater than the amount studied (Dam), it is a sign that the scale is tilted.
Emotion Management : As your E (Emotion) value increases, you may need to increase Ag (Agreement) to restore balance.
Self-reflection : You can check how well your current actions match your α(values).

학습 상황 : 공부한 양(Dam) 대비 집중력 저하(Dn)가 크면, 저울이 기울어져 있다는 신호.
감정 관리 : E (Emotion) 값이 커지면, Ag (Agreement)를 높여 균형을 회복할 필요가 있음.
자기 성찰 : α(가치관)를 기준으로, 현재 행동이 얼마나 일치하는지 점검 가능.

4. Academic Significance (학술적 의의)

Philosophical Implications : Formalizes the idea that humans, LLMs, and society as a whole seek balance on the “scale of reasoning.”

Mathematical Structure : Mitigates extremes through logarithmic transformation and provides an intuitive probability scale via the sigmoid function.

Applicability : Offers potential for interdisciplinary research in philosophy, psychology, AI ethics, politics, sociology, and data science.

철학적 의미: 인간과 LLM, 사회, 개인 모두가 “이치의 저울” 위에서 균형을 찾는 존재임을 수식화.

수학적 구조: 로그 변환으로 극단을 완화하고, 시그모이드로 직관적 확률 눈금을 제공.

응용 가능성: 철학, 심리학, 인공지능 윤리, 정치, 사회, 데이터 과학 등 다학제적 연구에 활용 가능성.

5. Final Greetings (끝 인사)

The **Scales of Reason** stand ready to weigh the weight of the counterweights.
Scholars, lay your dimensionless,
exquisitely refined counterweights upon the **Scales of Reason**.

I, Son Ho-Sung, am the original discoverer and author of the SEHE framework as described in this document, first disclosed on January 08, 2026.
Son Ho-Sung (*Registry Reference : 2026-SEHE-01-08, Republic of Korea*)

무게추의 무게를 저울 질 할 이치의 천칭이 준비 됐습니다.
학자님들의 손끝에서 다듬진 무차원화 된 아름다운 무게추를 이치의 천칭에 달아주세요.

본인 손호성은 2026년 1월 8일에 최초로 공개된 본 문서에 기술된 SEHE 프레임워크의 최초 발견자이자 저자입니다.
손호성 (*Registry Reference : 2026-SEHE-01-08, Republic of Korea*)

mail : maxsion1@gmail.com